

AN EMPIRICAL EVIDENCE OF CLUSTERED EARNINGS MANAGEMENT AND DEPENDENCE ON GOVERNANCE FACTORS IN SELECT LISTED FIRMS IN INDIA

DR. SOMNATH BANERJEE¹ & DR. SUBHASIS DASGUPTA²

¹Associate Professor, Department of Commerce and Management, St. Xavier's University, Kolkata, India

²Associate Professor, Praxis Business School, Kolkata, India

ABSTRACT

In this paper, we have computed the discretionary accruals which is the proxy for earnings management, for 527 firm-years using Jones' Model. Through background study, we have ascertained five independent variables which may affect the quantum of earnings management. Through a regression, we have found that bank debt and institutional holding have considerable influence on earnings management while audit quality is not so important. We have applied cluster analysis and have found clustered behaviour of firms in earnings management. Within the clusters, the goodness of fit of the regression also increases.

KEYWORDS: *An Empirical Evidence, Clustered Earnings Management, Dependence on Governance Factors & Firms in India*

Received: May 05, 2022; **Accepted:** May 25, 2022; **Published:** Jun 10, 2022; **Paper Id.:** IJECRJUN20229

1. INTRODUCTION

The literature in the earnings management area since Healy(1985) have extended our knowledge largely in understanding a) measurement models b) comparative analysis of efficiencies of different models in correctly measuring earnings management c) company specific causes or factors which affects earnings management and d) research on one of these areas in the perspective of different countries. (Holthausen *et al.*, 1995; Klein, 2002 and Alali, 2011) have established the implications of audit quality, executive compensation structure and executive compensation and its dependence on earnings levels. While most of the studies have centred around the scenario of the USA based firms and the European firms, the study on emerging economies like India is not in appreciable numbers in this area. It is well understood that the emerging economies are greatly different than US and European economies in terms of structure, regulatory framework, financial reporting styles and size of the firms. The need for a comprehensive study of earnings management and its contributories is of utmost importance in Indian context.

Mehran (1994) have studied the connection between the structure of performance based compensation and the level of earnings management. His research has thrown light on the area of whether the board is characterized by the dominance of the insiders than the outsiders and have then established the relationship of both these types with the ability of the managers to go for earnings management. His study is one of the pioneers in establishing a significant role of the structure of CEO compensation with the level of accruals on discretion. (Bergstresser and Philippon, 2006) have described the stock based compensation of the CEO and how it affects the entire process of earnings management. On the other hand (Haltahusen *et al.*, 1995) have explained some improvements of Healy(1985) model and have introduced a new method of computation of discretionary accruals. This paper also helped us to think about working on the most effective part of the data as at the extreme levels of income, when it is

too high or too low, the earnings management no more remains expected. So there must always be a part of the data where the earnings management values would be insignificant.

Our study is unique in the field of earnings management due to 1) it's comprehensive approach in Indian context. India is one of the biggest economy in the emerging markets. It is of utmost importance to examine the pattern of influences of different company specific factors on earnings management in India. Our study is built on a sample of 527 firms' years data of a wide cross section of Indian firms. With this we have tried to notice the ways in which earnings management is affected. We have analysed the influences of board composition, auditor's quality, financial leverage, CEO compensation through stock options and ownership patterns. For this part of the study, we have applied pooled OLS regression. 2) Taking cognizance of the findings of (Holthausen *et al.*, 1995), we have used the techniques of clustering and found it true that earnings management in India is a clustered scenario and the OLS regression within the significant clusters gives more robust results in explaining the dependence of earnings management on the company based factors. After clustering and selecting the significant cluster, the improvement in the quality of fit of the model is almost three times than before.

Selection of independent variables has been made out of vast area of previous literature and background of models, experiments and research which is explained in the next section. The entire study is divided into four sections. Section A is the introduction, section B is the background study, Section 3 is the model and methodology and Section 4 is the conclusion.

2. BACKGROUND OF THE STUDY AND RESEARCH DESIGN

2.1 Earnings Management and the Measurement

Earnings Management as already has been introduced, is the ways and means of earnings manipulation by the incentivized managers so as to maximize their pay offs by way of exercising the stock ownership plans which they are offered. In our research, we have applied Jones' Model to calculate Discretionary Accruals (DA) which is the proxy by which the earnings management is measured. In our research, we have worked with the absolute values of DA as the manipulation of earnings may be both upwards and downwards. The application and methodology part is explained in Section 3 of our discussion. The measurement of DA and the Jones Model is as follows:

The model starts with calculation of TA, the total accrual by the following equation:

$$TA_{i,t} = (\Delta CA_{i,t} - \Delta CL_{i,t} - \Delta Cash_{i,t} + \Delta STD_{i,t} - \Delta Dep_{i,t}) / A_{i,t-1}$$

After the TA has been calculated, it is divided by the last year's asset value. This is done for the purpose of normalization which is required to make the data comparable as per the model itself.

Once TA is calculated then TA is expressed as a function of three variables, namely Lagged Asset, Change in Revenue or Change in Sales and Property Plant and Equipment. All these variables have to be normalized using lagged assets. As we cannot divide the lagged asset itself by the same term, so in this model, the reciprocal of the lagged asset is taken.

As the next step, a regression is formed; taking TA as dependent variable and Reciprocal of Asset, Change in Revenue and Property Plant and Equipment as independent variables. In this regression the standardized coefficients are taken as all data are normalized. The coefficients of the three independent variables are then recorded for further

computation.

Then the actual values of Property Plant and Equipment, Change in Revenue and The reciprocal of the Lagged Asset are multiplied with the fitted coefficients. By doing this what we get is the Non Discretionary Accrual.

The difference between the Total Accrual and the Non Discretionary Accrual is the Discretionary Accrual.

The entire process is being presented in equation form as follows.

$$TA_{i,t} = \alpha_0 + \alpha_1 \times (1 / A_{i,t-1}) + \alpha_2 \times (\Delta REV_{i,t}) + \alpha_3 \times (PPE_{i,t}) + \varepsilon_{i,t}$$

$$\text{Estimated } NDA^2_{i,t} = \text{Estimated } \alpha_0 + \text{Estimated } \alpha_1 \times (1 / A_{i,t-1}) + \text{Estimated } \alpha_2 \times (\Delta REV_{i,t}) +$$

$$\text{Estimated } \alpha_3 \times (PPE_{i,t}) \dots \text{equation (iii)}$$

$$\text{Estimated } DA^2_{i,t} = TA_{i,t}^{\text{Estimated}} - NDA^2_{i,t} \text{ equation (iv)}$$

2.2 Bank Loan and its Influence on DA

In line with Mehran (1994), we also look into the basic premises that Long Term Bank Loan influences the process of construction of accruals. Watts and Zimmerman (1986) pointed out clearly that there is a tendency of manipulation of earnings management (upward) to gain reputation in debt market. This is most likely to happen practically too. In Indian Balance Sheets, it is evidenced that Bank Loan makes almost the entire long term corporate debts. Hence, in our research, we have taken Bank Loan as an independent variable. What we envisage is that it is possible that for a highly leveraged firm, the earnings management is just in the opposite direction than what Limas and his co-authors postulates, as we believe, for a highly leveraged firm, the financial reports would be of great concern to the bank and thus curbing the scope of managerial discretion. As we have already mentioned, we are attempting to predict the level of earnings management with artificial neural networks and hence at this moment we have refrained ourselves from looking into the matter of whether the relationship is positive or negative with the particular variable. We have to declare that bank loan goes into our multilayer perceptron model as an independent variable.

2.3 ESOP and its Relation with DA

One of the major factors behind earnings management, as has been explained by authors from time to time, is the stock ownership plans in the form of stock ownership. (Holthausen *et al.*, 1995 and Bergstresser & Philippon, 2006). The main reason of evolution of this specific branch of research basically evolves from the concept of accounting choices of Watts and Zimmerman (1990). In Indian listed firms, we have gone through a wide cross section of firms with ESOP as a component of executive compensation and have considered ESOP as a significant variable that influences earnings management. We have not tried to look into whether incentivized managers would go for an income inflating or an income deflating earnings management depending on his choice based upon the influence of the situation, rather, we have concentrated on the issue of using this as a predictor for the actual level of earnings management as proxied by DA.

2.4 Audit Quality, Board Character and Earnings Management

There has been close association in between the audit quality and earnings management Klein (2002). It is evidenced in the literature too, that a high quality audit reduces the flexibility of managers in manipulating the accounts. Klein, has established the fact that an Audit Committee which is dominated by independent directors, gives less chance to the business unit level managers to inflate or deflate reported income. Audit fee and discretionary accruals have relations too

Alali (2011). This points out to the very fact that quality of auditor has an important impact on the level of earnings management. In our research, we have considered two classes of audit quality i.e. high quality audit (when the auditor belongs to big four) with certain exceptions and the other category when the auditor is not among the big four. We have not paid attention towards audit committee as such, rather we have used the auditor's reputation as a proxy for the quality of audit and have used this as an independent variable in our research.

We have taken the Board characteristic in our model as an independent variable too much in accordance with Klein (2002). The only point which needs clarification is that we have restricted ourselves within the percentage of independent directors in the board. We expect that a higher percentage of independent directors in the board will for sure, make the manipulation in financial figures challenging for the managers. There should be an inversivity of relationship here but as we have already mentioned that we have put a serious effort in predicting the level of earnings management and not in the direction of association. In our neural network, both independent director and quality of auditor is fed as an independent variable.

2.5 Institutional Ownership and Earnings Management

Institutional Ownership, Block holding patterns and relationship investing (Khan and Mather 2013) have established the relationship of institutional holding and earnings management. Klein (2002) have mentioned about 'relationship investment' which is the block holding by big investors. We have considered the presence of institutional holding as an important variable in our study as we live, much in line with the background work that an institution holding a larger part of the ownership will always try to exercise his control over the company in his own interest itself, which may be in way of keeping a nominee director in the board or by some other methods which are not the focus of discussion in this paper. Whatsoever is the way of exercising control by the block holders, their presence would have a definitive influence on the discretion of managers in manipulation of earnings to be reported. We have included the institutional ownership pattern as a percentage in the list of our independent variables.

In Section 3, we have elaborately discussed the use of all these variables and the exact nature or pattern in their influence on discretionary accruals.

3. RESEARCH METHODS

This section describes the sample data which has been used and the various methods, analytical tools etc which have been applied to conduct this research. This section is divided into two distinct parts. Section 3A lays down the entire process of calculating DA and section 3B explains the analytical methods which have been applied to extract the relationship of DA with those firm specific factors. As already mentioned, section 3B shows the nature of relationship which five different factors have with the level of earnings management of firms in India.

3A Measuring Discretionary Accrual (DA)

This study have been conducted with 527 firm year's data across a wide cross section of Indian firms. The data has again two dataset with distinctive properties. Out of the entire data, one part consists of all the firms which are there in the leading indices of India (SENSEX, NIFTY) and hence this part of the sample is a non probabilistic sample. The other part of the sample data is the collection of the other category of Indian firms which are not included in the major indices of the country. The reason for aggregating these two parts is to bring a reasonable broad base to the study and to include a larger

cross section of Indian firms under study. The other part of the sample includes companies with substantial trading volume in the market in a ranked order which is available from the stock exchanges itself. The following section shows the fragments of the sample in greater details. The variables which are required for the model are Change in Current Liabilities, Current Assets, Changes in Depreciation, Changes in Cash from operations and changes in long term debt which is falling due in immediate future. Due to unavailability of data for all companies in this item, it has been kept aside for the time being. The snapshot of the entire sample is presented in Table 3.1 and figure 3.2

Table 3.1: Description of Companies in Sensex & Nifty				
Item	Minimum	Maximum	Mean	Median
Asset	639.41	295140	46385	25117.71
Sales	212.74	414919	38922	20969.80

Table 3.2: Description of other Companies

Item	Minimum	Maximum	Mean	Median
ASSET	349.59	99326.8	9592.29	3629.09
SALES	12.01	197744	5352.13	3584.55

As per the Jones Model, for both the samples, Total Accrual (TA) has been calculated from the model function itself where the independent variables are changes in PPE, REV and inverse of lagged asset (Details of the terms and model specifications are discussed in section 2.1). TA is then expressed as a regression function with these variables. Then the regression parameters are multiplied with the actual values which gives NDA. DA is the difference of TA and NDA. The TA as calculated for big and medium firms are shown in Table 3.3

Table 3.3 Regression for NDA from Total Accruals (TA)

$NDA_{i,t} = .010 \times (1 / A_{i,t-1}) - .017 \times (\Delta REV_{i,t}) + .408 \times (PPE_{i,t}) \dots (\text{medium})$
$NDA_{i,t} = -.305 \times (1 / A_{i,t-1}) + .007 \times (\Delta Sales_{i,t}) + -.007 \times (PPE_{i,t}) \dots (\text{big})$

As already described in the model, DA is the differences between TA and NDA as per the model in use. The data before putting into experimentation, had been cleaned for outliers in SPSS using Boxplot method. This is the first stage outliers. A second stage outlier elimination process, for the reason explained in section 3B, is also elaborated in the next section.

In our study, we have taken the absolute values of DA by using abs function in Excel as we have already stated that the direction of DA is not relevant in this research and it is the magnitude which is important.

3B Relationship of DA with Firm Specific Variables

In this section the relationship of DA with five firm specific variables, (which have been elaborately discussed in Section 2) has been explored by the help of machine learning techniques using Python and by simultaneous operation of conventional statistical techniques using SPSS. The five dependent variables as have been used in the study are BIGAUDR (explains whether the auditor is Big Four), IHPP (explains institutional holding pattern percentage), TDEBTN (Total Debt), ESOP (Employee Stock Ownership Plan), INDRCR (Proportion of Independent Director in Board). Initially all variables are normalized by lagged values of Assets as per Jones Model.

The dependent variable is the absolute values of all DA as measured in Section 3A and they have been combined to make the set of DAMOD.

It is drawn heavily from past studies (as described in Section 2) that there is a range bound behaviour of earnings management by managers. When the income is too high or too low, then as evidenced, the earnings management practices are minimum. We have not tried to map earnings management with earnings bands or levels to draw inferences, rather we have tried to apply non linearity assumption of the relationship (which we have established later) and have established the presence of clusters in a data of this nature. The description of the dataset which is put under research is visible in Table 3.4

Table 3.4

	N	Minimum	Maximum	Mean	Std. Deviation
TDEBTN	527	.0000	3.7514	.150459	.2759299
DAMOD	527	.0001	.9331	.126750	.1228493
IHPP	527	.0281	.6332	.285235	.1340087
INDRCTR	527	.0000	.9000	.490559	.2448973
ESOP	527	0	1	.33	.471
BIGAUDR	527	0	1	.58	.494
Valid N (listwise)	527				

3.B.1 Linearity Assumption and Regression

The initial pooled OLS regression which was attempted has given the result which is presented in Table 3.5

Table 3.5 Summary Output of Initial Regression before clustering ***

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.365 ^a	.13125	.1149184	

The above results prompted us to predict the existence of clusters in the data and then clustering was attempted to identify the existence of valid clusters.

3.B.2 Existence of Clusters and OLS Regression in Significant Clusters

A. To see if data were clustered, a dimension reduction technique was required to either project or embed the higher dimension to lower dimensions. TSNE method was put forward by Maaten and Hinton (2008) which can embed higher dimensional data into lower dimensions while trying to maintain the distribution of data as seen in the high dimension to the low dimension. Since the total number of variables was 6, TSNE plot was considered to be appropriate to see the presence of clusters in the dataset. TSNE plot clearly shows the presence of clusters in the dataset and hence a single regression line would not be in a position to capture the pattern in a plausible manner. This is visible in Figure 3.1

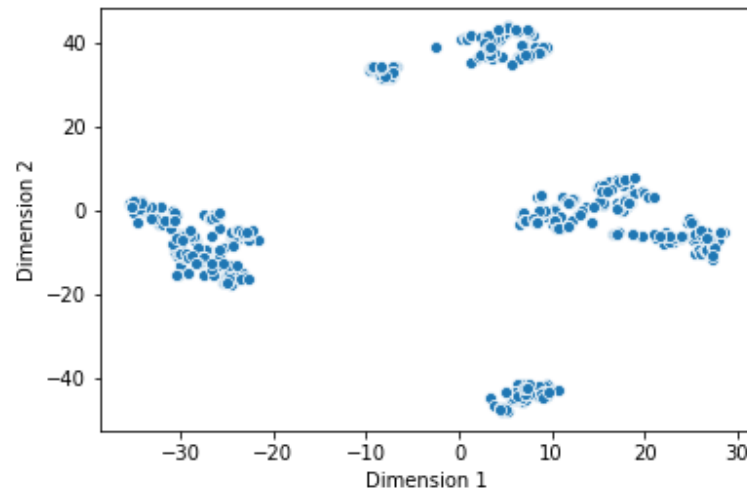


Figure 3.1: Clusters (TSNE).

To capture the clusters, Expectation Maximization Clustering (EM Clustering) (Celeux and Govaert, 1992) method was employed. EM clustering is a soft clustering technique which associates each data point to every cluster so that uncertainty of data clustering can be understood objectively, a feature not available in hard clustering technique such as KMeans clustering. Moreover, EM being a statistical method that calculates the likelihood values, Akaike Information Criteria (AIC) (Sakamoto et al., 1986) and Bayesian Information Criteria (BIC) (Watanabe, 2013) values can be calculated quite easily for EM algorithm. A good model is that which has lowest AIC or BIC scores. BIC being more conservative than AIC, BIC score was considered for evaluating the cluster solutions. When EM clustering was run on the dataset for different clustering solutions, it was seen that the lowest BIC value was appearing at 12 cluster solution as shown in Figure 3.2. However, at the 5 cluster solution, there was a considerable amount of reduction in the BIC score and the score was quite similar to that of the 12 cluster solution. Moreover, the TSNE plot also showed 5 clusters in the dataset. Hence, the next task was to validate the 5 cluster solution using a validation technique. Silhouette score is a popular cluster validation technique and when the silhouette score was found out for the 5 cluster solution, it came out to be 0.60 which signified that the solution was reasonably valid and the clusters were well separated.

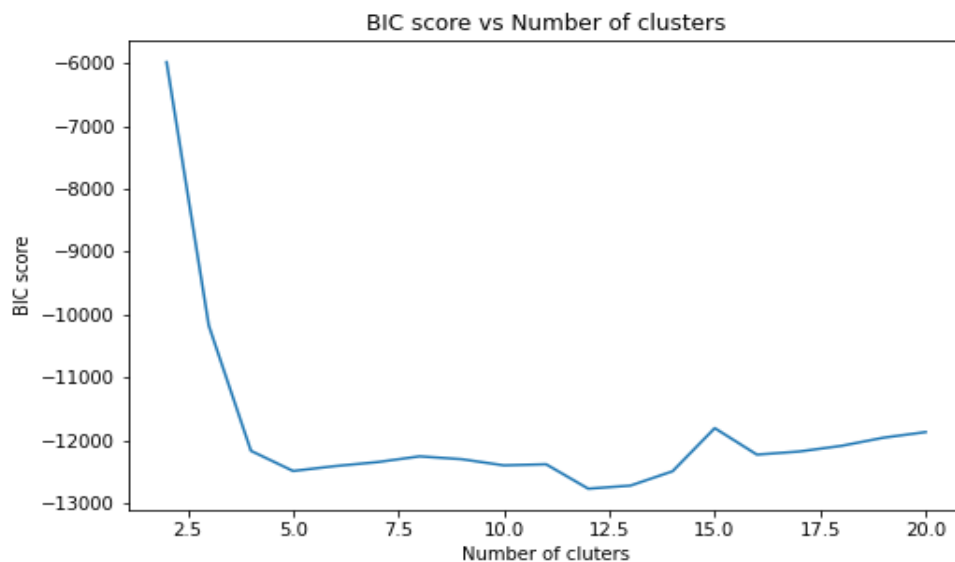


Figure 3.2

The next task was to remove data points which were outliers. To achieve this objective, Local Outlier Factor method was administered on the dataset. Local Outlier Factor is a popular nonparametric method of outlier detection and using this method, outliers were detected and removed from the dataset. In doing so, 24 data points were removed as outliers. With that removal, the 5th cluster was also removed. Afterward, linear regression was run in each cluster considering DAMOD as the dependent variable. The R-sq values are shown in Table 3.5. Clearly, no significant dependency of DAMOD on other variables could be seen in cluster 0 and cluster 1. The variance explained by the model was only 4% in cluster 0 and 0.5% in cluster 1. Only for cluster 2 and cluster 3, some amount of dependency could be seen.

Table 3.5 R Squares in Different Clusters

Cluster No	0	1	2	3
R Sq		.040	.005	.160

B. Cluster 2 and 3 were clubbed to form the new dataset and the regression was performed with SPSS which proved substantial enhancement in the result which is presented in Table 3.6

Table 3.6: The Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
2	.561 ^a	.315	.305 ***	.0762114

Table 3.7: The Coefficients

Variables	TDEBTN	IHPP	INDRCTR	BIGAUDR	ESOP
Values.	148 -	107	030.	068	Insignificant

Nonlinear regression using Machine Learning: For this analysis also, dataset having cluster number 2 and 3 were considered. Since nonlinear methods have a tendency to remember the data leading to very high level of performance on training data but poor performance on test data (known as model over fitting), the dataset was partitioned into 2 parts (i.e. **training_set** and **test_set**) with 60:40 ratio so that the models can be trained on the training set and their performances can be evaluated on the test set. For the said purpose, three powerful non-linear ML methods were used to predict the dependent variable DAMOD. The three models used were Random Forest, Xgboost and Light GBM. Since performance of these models are sensitive to the choice of hyperparameters, the hyperparameters were chosen based on 5 fold cross validation using evolutionary search method for each model. Afterward, all the three models were trained using the training set and then they were applied on the test set to see the performance. RMSE was used to measure the performance of these models.

The RMSE scores are:

Table 3.8

Model	RMSE (based on Test Dataset)
Random Forest	0.0693
Xgboost	0.0762
LightGBM	0.074

When the similar approach was run with linear regression, RMSE was found to be 0.079 and hence, it could be seen that nonlinear regression with Random Forest provided reasonably higher level of performance compared to linear

regression.

4. CONCLUSIONS

- Earnings Management exists across a wide cross section of listed firms in India which is mostly influenced by the quality of audit, institutional holding, Board characteristics and level of debt. ESOP, on the other hand does not play any significant role in determination of the level of earnings management in India. This is in consonance with our statement in Section A, where we have mentioned that the economy of emerging markets have distinguished characteristics like this, where we may have entirely a different scenario in the US and European economies.
- Bank Debt shows a significant positive correlation with earnings management in India. It means that to secure debt or extension of rolling period, the companies engage in inflating the reported numbers.
- Auditor's quality on an overall basis is not very prominent in controlling earnings management in India. It can also be concluded that big firms appoint big auditors. Mostly firms which are big have more analyst following and they have big auditors. These firms, on the other hand, have an urgency to maintain its stock prices with an upward bias. Hence they may engage in inflating reported numbers. Audit quality may not be of very high standard.
- Block holders or institutional holders exercise a greater control on the management and they have a significant inverse relation with the earnings management level.
- Out of the entire economy, earnings management practices are clustered. Some clusters are insignificant where the earnings management and the contributing factors have no relationship at all. But there are significant clusters, which have significant relationship of earnings management with the contributing variables.
- The cluster based regressions improves the results of the tests by almost three times as expected and hypnotized in Section 1. By application of various models, the values of earnings management which we calculate, some part of it is just sporadic and originating because of application of models. Cluster based approaches may eradicate them and may prove to be a better way of thinking about this issue.

REFERENCES

1. Alali, F. (2011), 'Audit fees and discretionary accrual compensation structure effect', *Managerial Auditing Journal*, Vol. 26 No. 2, pp. 90-113.
2. Bergstresser, D and Philippon, T. (2006) 'CEO compensation and earnings management', *Journal of Financial Economics*, Vol. 80 No. 3, pp. 511-529
3. Celeux G and Govaert, G. (1992), 'A Classification EM algorithm for clustering and two stochastic versions.', *Journal of Classification*, Vol. 14 No. 3, pp.315-332
4. Healy, P. (1985) 'The effect of bonus schemes on accounting decisions', *Journal of Accounting and Economics*, Vol. 7 No.1, pp.85-107
5. Holthausen, R., Larcker, D and Sloan, R. (1995) 'Annual bonus schemes and the manipulation of earnings', *Journal of Accounting and Economics*, Vol. 19 No. 1, pp.29-74
6. Jones, J. (1991) 'Earnings management during import relief restrictions', *Journal of Accounting Research*, Vol. 29 No. 2, pp.193-228

7. Khan, A and Mather, P. (2013), 'The value of executive director share ownership and discretionary accruals', *Accounting Research Journal*, Vol. 26 No. 1, pp.35-55
8. Klein, A. (2002) 'Audit committee, board characteristics and earnings management', *Journal of Accounting and Economics*, Vol. 33 No. 3, pp.375-400
9. Maaten, L. V. D and Hinton, G. (2008). 'Visualizing data using t-SNE', *Journal of machine learning research*, Vol. 9 No. 85),pp. 2579-2605
10. Mehran, H. (1995) 'Executive compensation structure, ownership and firm performance', *Journal of Financial Economics*, Vol. 38 No. 2, pp.193-228
11. Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. Dordrecht, The Netherlands: D. Reidel, 81.
12. Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research*, 14(Mar), 867-897.
13. Watts, R. L and Zimmerman, J. L. (1990), 'Positive Accounting Theory', *Accounting Review*, Vol. 65 No. 1,pp.131-156
14. Joseph, Benny. "An Empirical Study on Knowledge Management in Higher Educational Institutes: A Case Study of Christ Campus, Rajkot." *International Journal of Humanities and Social Sciences (IJHSS)* 6.3 (2017): 111-120.
15. Sreelalitha, K. G., and S. Shankarappa. "Job Stress and Performance: an Empirical Study with Special Reference to Teachers Working in Higher Educational Institutions in Chamarajanagar District." *International Journal of Business General Management (IJBGM)* ISSN (P): 2319-2267; ISSN (E): 2319-2275 Vol. 10, Issue 2, Jul-Dec 2021; 17-24